

DOI: 10.20040/j.cnki.1000-7709.2023.20221474

# 基于时变特征矩阵的相似流域识别方法研究

李珂<sup>1,2,3</sup>, 阚光远<sup>1,2,3</sup>, 何晓燕<sup>1,2,3</sup>, 祝冰洁<sup>4</sup>

(1. 流域水循环模拟与调控国家重点实验室, 北京 100038; 2. 中国水利水电科学研究院, 北京 100038; 3. 水利部防洪抗旱减灾工程技术研究中心, 北京 100038; 4. 河南省水文水资源局, 河南 郑州 450003)

**摘要:** 为了提升无径流资料区水文预报相似流域识别的可靠性, 提出了一种相似流域识别方法, 该方法通过引入流域时变特征信息, 构建流域时变特征矩阵及其展布方法获取流域特征向量, 并利用经典 K-means 聚类方法实现相似流域识别, 并以江西省 17 个天然闭合流域为例, 结合新安江模型, 开展特征向量/矩阵构建、相似流域聚类识别、模型参数移用、水文预报结果对比分析等一系列研究。结果表明, 提出的时变特征矩阵用于无径流资料区水文预报能取得较好的应用效果, 与传统的仅基于时不变下垫面信息的相似流域识别方法相比, 预报精度得到提升; 基于时变综合要素获得的相似分组取得了最高的预报精度, 基于时变气象要素获得的相似分组预报精度优于仅基于时不变下垫面要素获得的相似分组; 与基于单一要素相似流域识别方法相比, 基于综合要素获得的相似分组改善了参数移用的方向性问题, 但尚不能完全消除方向性影响。

**关键词:** 相似流域; 时变特征矩阵; 无资料地区; 参数移用; 新安江模型

**中图分类号:** TV12 **文献标志码:** A **文章编号:** 1000-7709(2023)06-0013-04

## 1 引言

无径流资料区(简称“无资料区”)水文预报一直是水文学领域的重难点问题。目前, 已提出了大量无资料区水文预报的新思路、新方法, 其中参数区域化是应用最多的实用方法。该方法的核心思想是将有资料区水文模型参数移用到与该区域相似的无资料区, 以便利用移用后的参数开展水文预报工作。合理地将模型参数外推至无资料流域是参数区域化方法面临的难题。当前常用的方法是利用流域下垫面特征构建水文相似因子, 基于此对流域进行分类进而实现模型参数向无资料流域的移用<sup>[1,2]</sup>。随着研究深入, 发现仅考虑下垫面因子识别水文相似性, 并不能完整揭示流域水文现象的本质<sup>[3]</sup>, 因此将气象因素或水文要素引入相似流域识别研究中, 取得了较好的应用效果<sup>[4]</sup>。无资料区水文预报研究大多基于流域下垫面特征构建相似因子, 很少考虑流域气象特征因

素。此外, 对于流域相似性识别的准则大都采用均值处理后的特征值, 未考虑特征值动态变化过程, 造成了大量信息损失, 使相似性识别结果出现误差。因此, 本文在传统下垫面要素的基础上, 引入具有时变特性的气象特征构建特征矩阵, 基于此开展相似流域识别研究, 以期改善无资料区相似流域识别和水文预报效果。

## 2 研究区域与数据

### 2.1 研究区域概况

江西省地跨东经 113°35′~118°29′、北纬 24°29′~30°05′, 省内降雨丰沛, 汛期较多, 空间分布不均。选取江西省内瑞金、羊信江、翰林桥、信丰、赛塘、新田、沙子岭、南丰、南城、弋阳、渡峰坑、深渡、香屯、虎山、婺源、德兴、万家埠 17 个天然闭合流域作为研究流域(均匀分布于省内), 流域面积在 486~8 600 km<sup>2</sup> 之间。

**收稿日期:** 2022-07-18, **修回日期:** 2022-08-19

**基金项目:** 国家重点研发计划(2019YFC1510605); 光合基金 A 类 20210701(ghfund202107011171); 中国水利水电科学研究院十四五“五大人才”计划(JZ0199A032021); 中国水科院减灾中心“基础研究型”科技创新人才项目; 北京师范大学北京市重点实验室开放基金(HYD2020OF02)

**作者简介:** 李珂(1996-), 女, 硕士研究生, 研究方向为水文预报, E-mail: 1850195367@qq.com

**通讯作者:** 何晓燕(1972-), 女, 博士、正高级工程师, 研究方向为洪水预报与调度、水库汛期运行水位动态控制等, E-mail: hexy@iwahr.com

## 2.2 数据来源与预处理

本文相似流域识别驱动数据涉及下垫面特征和气象特征。其中,下垫面特征包括流域地形特征、土地利用/覆盖、归一化植被指数(NDVI);气象特征包括实测降雨序列、长波辐射等。对各特征值的时间范围取交集,确定共有数据范围为2015年。数据类型、数据来源等信息见表1。基于所选数据,分别构建下垫面特征向量、气象特征矩阵、综合特征矩阵。采用2015~2019年场次洪水进行参数率定。

表 1 选取的数据集

Tab.1 Selected data sets

数据类型	数据时空分辨率	数据来源
实测降雨系列	1 h、台站	江西省水文局
气象特征	1 h、0.5°×0.5°网格	GLDAS 数据集
实测径流系列	1 h、台站	江西省水文局
高程数据	时不变、90 m 网格	ASTER GDEM V2 数据集
土地利用/覆盖	时不变、300 m 网格	中科院地理所
土壤质地	时不变、300 m 网格	欧空局
归一化植被指数(NDVI)	时不变、250 m 网格	MODIS MOD13Q1 数据集

## 3 相似流域识别方法

### 3.1 参数率定

为了便于验证结果,本文涉及的流域均为有资料流域,利用SCE-UA算法对新安江模型开展参数率定,获得最优模型参数。考虑到纳什效率系数(NSE)能综合反映水文过程线的拟合程度,因此采用纳什效率系数作为预报精度的评定指标。

### 3.2 相似流域识别

#### 3.2.1 构建特征矩阵与特征向量

传统相似准则的构建是将各类特征在时间上视为常量即单值化处理,将单值化后的特征拼接成一维特征向量。引入特征的时变特性,在一维特征向量的基础上将特征值在时间尺度上进行扩展形成时变特征矩阵,特征矩阵的构建方式见图1。

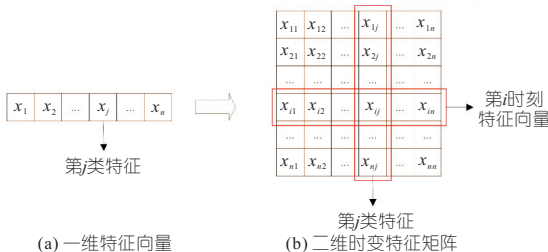


图 1 一维特征向量与二维特征矩阵

Fig.1 One-dimensional feature vector and two-dimensional feature matrix

#### 3.2.2 特征矩阵展布方法

为避免特征值量纲对相似流域识别带来不良影响,使用极值标准化方法对数据表进行归一化处理。采用经典K-means方法,基于特征向量进

行相似流域识别。由于特征矩阵无法满足聚类算法输入模式的要求,因此提出了特征矩阵展布方法,即将每一列作为一个元素,多列进行拼接,组成一个多元素的向量将特征矩阵转换为特征向量。

### 3.3 参数移用

本文主要移用产流蒸散发参数,汇流参数由无资料区参数推演方法获得<sup>[5]</sup>。为了研究基于不同准则进行相似流域识别的效果,开展三套准则下的相似流域识别研究。对于三套准则,分别利用K-means算法进行聚类,将17个研究流域划分为 $N_k$ 个相似流域组( $k$ 表示相似流域聚类基于第 $k$ 套准则, $k=1,2,3$ )。任取相似流域组内的一个流域为参证流域,组内其他流域为目标流域,开展参数移用、水文预报及精度评定。

## 4 结果与讨论

### 4.1 相似流域识别

#### 4.1.1 识别准则的获取

(1)准则1。下垫面特征向量。由流域面积、平均高程、平均坡度、河网密度、平均地形指数、土地利用/覆盖、土壤质地、归一化植被指数构成。其中,土壤质地、植被密度、土地利用/覆盖等特征主要影响流域产流过程;其余特征主要影响流域汇流过程。由于下垫面特征年内变化不明显,故下垫面特征采用非时变数据,将下垫面特征值连接构成一维特征向量的形式。

(2)准则2。气象特征矩阵。由实测降雨与长波辐射、短波辐射、风速、大气温度、比湿、地表气压、下行长波辐射、下行短波辐射构成。其中,降雨序列主要对流域的产流起作用,其余气象特征主要影响流域蒸散发过程。

(3)准则3。综合特征矩阵。基于气象特征矩阵,将下垫面特征向量在每一时刻作为常量进行扩张形成综合特征矩阵。

#### 4.1.2 模型参数率定

基于SCE-UA算法开展新安江模型参数率定,选取洪水70%场次用于参数率定,30%场次用于验证,率定期与验证期的纳什效率系数箱形图见图2。由图2可知,目标流域率定期和验证期水文模拟的纳什效率系数均在0.6以上,表明率定获得的参数在该地区应用效果较好。

#### 4.1.3 相似流域识别

将归一化后的特征矩阵或特征向量作为聚类准则,使用矩阵展布方法和K-means聚类方法进行相似流域识别。计算聚类数目 $K$ 取不同值时的轮廓系数,当 $K=3$ 时,各聚类准则下的轮廓系

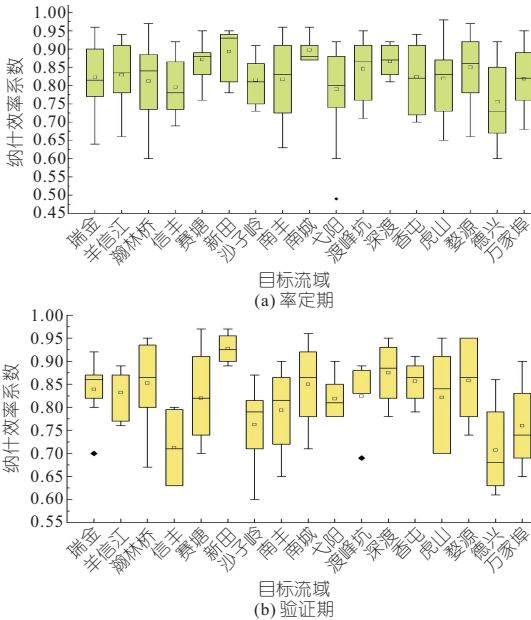


图 2 率定期、验证期洪水纳什效率系数箱型图

Fig. 2 Nash-sutcliffe efficiency coefficient box diagram of flood in calibration and validation period

数均取得最大值,表明分类效果此时达到最优。三套准则下获得的相似流域分组详情见表 2。

表 2 不同聚类准则下获取的相似流域识别结果

Tab. 2 Similar watershed identification results obtained under different clustering criteria

识别准则	相似流域组 1	相似流域组 2	相似流域组 3	识别准则	相似流域组 1	相似流域组 2	相似流域组 3
时不变下垫面特征向量	渡峰坑 南城 南丰 沙子岭 深渡 香屯 信丰 羊信江	瀚林桥 瑞金 赛塘 万家埠 弋阳	德兴 虎山 婺源 新田 弋阳	时变综合特征矩阵	德兴 瑞金 沙子岭 婺源 羊信江	渡峰坑 虎山 南城 南丰 深渡 香屯 新田 弋阳	赛塘 万家埠 信丰
时变气象特征矩阵	德兴 瑞金 沙子岭 羊信江	瀚林桥 南城 南丰 万家埠 信丰	渡峰坑 虎山 赛塘 深渡 婺源 香屯 新田				

4.2 结果分析

统计目标流域移用参证流域参数后水文预报纳什效率系数 ≥ 0.9 的洪水场次占比称为甲级预报合格率,纳什效率系数 ≥ 0.5 的洪水场次占比称为合格率。甲级预报合格率与合格率越高,参数移用效果越好。

以同一相似流域组内流域为例,研究参数移用方向性问题,具体方法为相互移用相似流域 A、B 的水文模型参数(流域 A、B 属于同一个相似流域组),统计对应的模拟结果合格率,并计算参数移用误差差值的绝对值(简称为方向值),当该值大于 20%,则说明参数移用存在明显的方向性问题,方向值越大参数移用方向性问题越明显。

4.2.1 不同准则前提下参数移用预报精度分析

针对三种不同聚类准则识别获得的相似流域组进行组内参数移用,统计水文预报合格率,结果见图 3(a)。由图 3(a)可知,综合特征矩阵合格率范围在 0.6~0.8 之间,均值较高,总体分布集中,数据差异性较小;气象特征矩阵的合格率均值次之,总体分布离散;下垫面特征向量的合格率均值最低,总体分布离散,数据差异性大。由此可知,通过引入时变气象因素,考虑了降雨强度等对洪水预报起关键性作用的因子,能显著提升水文预报的精度。

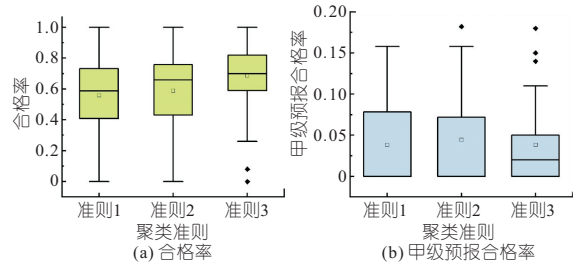


图 3 不同聚类准则平均合格率与甲级预报合格率分布

Fig. 3 Average qualified rate of different clustering criteria and qualified rate of grade A forecast

统计分析三种不同聚类准则下相似流域分组参数移用后水文预报的甲级预报合格率,结果见图 3(b)。由图 3(b)可知,三种不同聚类准则下甲级预报合格率均值均较接近。综合考虑合格率与甲级预报合格率,基于综合特征矩阵进行相似流域识别效果最好。

统计各目标流域的合格率均值与甲级预报合格率均值,结果见图 4。由图 4(a)可知,综合特征矩阵准则下各个研究流域的水文预报结果合格率均值整体较高,基于气象特征矩阵的水文相似性识别结果平均预报合格率次之,基于下垫面特征向量的水文相似识别的平均预报合格率相对较差。

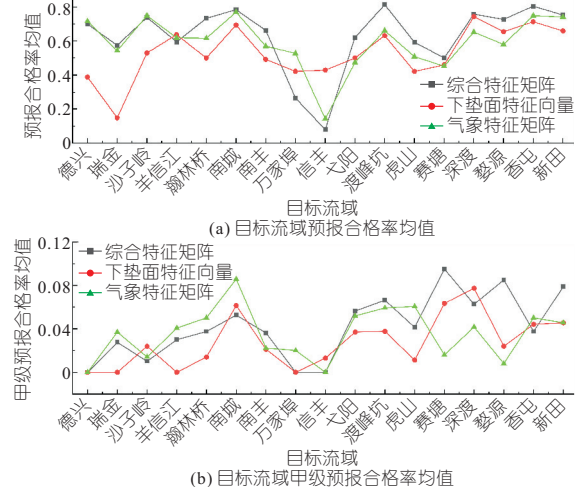


图 4 目标流域预报合格率均值、甲级预报合格率均值

Fig. 4 Average qualified rate of target basin forecast and average qualified rate of grade A forecast

由图 4(b)可知,综合特征矩阵和气象特征矩阵中各个研究流域的水文预报结果的平均甲级预报合格率各有优劣。下垫面特征向量的水文相似识别的平均甲级预报合格率相对最差。

#### 4.2.2 参数移用方向性问题分析

分别统计三种识别准则中的方向值见图 5。由图 5 可知,下垫面特征向量和气象特征矩阵相似流域组的方向值均值较接近,而综合特征矩阵相似流域组的方向值更小,且分布更集中。说明气象特征矩阵和下垫面特征相似流域组出现方向性问题相对较明显,综合特征矩阵相似流域组在方向问题方面表现更优。

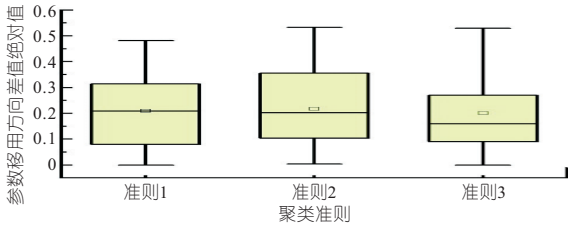


图 5 不同聚类准则参数方向误差统计

Fig. 5 Directional error statistics of different clustering criterion parameters

## 5 结论

a. 基于特征矩阵的相似流域识别方法可用于无资料区水文预报。与传统的仅基于下垫面信息

的相似流域参数移用方法相比,预报精度得到提升。

b. 基于特征矩阵聚类准则下,相似流域水文分组的参数移用结果整体上表现良好。其中,基于综合特征矩阵聚类准则下水文相似分组的参数移用效果最好,基于气象特征矩阵聚类准则下水文相似分组的参数移用效果次之,基于下垫面特征向量聚类准则下水文相似分组的参数移用效果相对较差。

c. 综合特征矩阵聚类准则相比单一因素聚类准则下的水文相似分组在参数移用方向性问题上表现更优,但方向性问题的彻底解决仍需进一步深入研究。

#### 参考文献:

[1] 吴国群,甘升伟,刘金涛. 基于模糊聚类法的无资料流域水文过程模拟[J]. 水电能源科学, 2018,36(10):13-16.

[2] 金倩芳,周建中,严冬,等. TOPMODEL 模型在长江上游无资料地区的水文模拟[J]. 水电能源科学, 2018,36(9):44-47.

[3] 刘金涛,宋慧卿,张行南,等. 新安江模型理论研究的进展与探讨[J]. 水文, 2014,34(1):1-6.

[4] 李晨涛. 基于主成分聚类的本溪无资料地区中小河流参数移植研究[J]. 水利规划与设计, 2021(11):41-45,69.

[5] 李致家,梁珂,阚光远,等. 新安江模型中河网汇流参数  $C_s$  的一种计算方法[J]. 水科学进展, 2016,27(5):652-661.

## Research on Similar Basin Recognition Method Based on Time-varying Feature Matrix

LI Ke<sup>1,2,3</sup>, KAN Guang-yuan<sup>1,2,3</sup>, HE Xiao-yan<sup>1,2,3</sup>, ZHU Bing-jie<sup>4</sup>

1. State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, Beijing 100038, China;
2. China Institute of Water Resources and Hydropower Research, Beijing 100038, China;
3. Research Center on Flood and Drought Disaster Reduction of the Ministry of Water Resources, Beijing 100038, China;
4. Henan Bureau of Hydrology and Water Resources, Zhengzhou 450003, China)

**Abstract:** A similar basin identification method was proposed to improve the reliability of similar basin identification for hydrological forecasting in ungauged basin. The method introduced the basin time-varying feature information and constructed a time-varying feature matrix (TVFM). A dimension reduction method for the TVFM was proposed to obtain the basin feature vector. The classical K-means clustering method was adopted to implement the identification of the similar basins. Seventeen natural catchments in Jiangxi Province were selected and the Xin'anjiang model was applied in this research. A series of studies were carried out which includes feature vector and feature matrix construction, similar basin identification, parameter transfer, and comparison & analysis of hydrological forecasting results. The following conclusions can be drawn. The proposed TVFM can achieve better forecasting results in ungauged basins. Compared with the traditional similar basin identification method based on time-invariant underlying surface information, the prediction accuracy is improved. The similar basin identification based on time-varying comprehensive features generates the best predictions, and the prediction accuracy based on time-varying meteorological features is better than that based only on time-invariant underlying surface features. Comparison with the single factor methods, the directional problem of the parameter transfer can be partially solved by the proposed comprehensive method. Unfortunately, the directional problem cannot be completely overcome until recently.

**Key words:** similar basins; time-varying feature matrix; ungauged basin; parameter transfer; Xin'anjiang model